**CS410 Final Project Report**

**Part 1 Overview of the function of the code:**

In the final project, we choose to implement a basic Expert Search System. In this system, when a user enters a specific field or area as a query, it will return relevant experts with relevant coefficients. Therefore, for this project, the code consists of three parts: The first part is to write a crawler to crawl data from Google Scholar, the second part is to generate relative results from these crawled data by utilizing some packages. The third part is to implement a basic front end with search queries and connect the data with the front end.

**Part 2 How the software is implemented:**

As mentioned in the Part 1, the codes mainly consist of three major parts: 1. Crawling data 2. Utilizing the elastic search package 3. Connecting the data with front end

I. Crawling Data

The basic codes of crawling data from Google Scholar is mainly inspired from the following link called scholar.py: <https://github.com/ckreibich/scholar.py>

With the help of the crawling codes, we can easily crawl much information from a specific professor. It is difficult to convert the papers of pdf files into ASCII text perfectly. Therefore, we choose to crawl the title and abstract instead. All the details are listed in “scholar.py” folder.

II. Utilizing the elastic search package

By utilizing the inverted-index from the elastic search package, we can easily train the crawled data of titles and abstracts and retrieve the relevant topics of these papers. All the details are listed in “json.py” file.

III. Connecting the data with the front end

The last part is to implement front end and connect it with the data, and the codes are listed in “searchPage.html”.

**Part 3. Usage of the website:**

To use this software, we should:

1. First install the elasticsearch package by typing “sudo pip3 install elasticsearch” in the command
2. Then we execute the elasticsearch api by typing “elasticsearch/bin/elasticsearch” on the terminal.
3. Next, we run the json.py by typing “python3 json.py” in the command.
4. Lastly, we add the Allow-Control-Allow-Origin plugin and open the searchPage.html in Chrome browser to start the software.

**Part 4 Work distribution:**

Zhuoran Li (netid: zli104): Progress report and crawling data from Google Scholar

Qingkang Zhang ( netid: qzhang72) : Implementing the back end

Qixuan Li (netid :qli44) : Implementing the front end